CLAIMS

What is claimed is:

Sub
a²

1.   A method for generating cross-references among categories in a knowledge base, said method comprising the steps of:

extracting, from a plurality of documents, a plurality of themes, wherein a theme identifies subject matter contained in a corresponding document;

generating a plurality of scores such that each score identifies a relative theme strength among theme pairs of said themes extracted from said documents, said theme strength reflects the amount of subject matter contained in a document for a corresponding theme relative to other themes in said document;

selecting theme pairs based on said scores;

selecting category pairs in said knowledge base by mapping said themes of said theme pairs selected to corresponding categories of said knowledge base; and

generating a cross reference in said knowledge base between categories of said category pairs, wherein said cross reference identifies an association between said category pairs.

2.   The method as set forth in claim 1, wherein the step of generating a plurality of scores comprising the steps of:

generating a matrix comprising a plurality of columns and rows to form a plurality of entries, wherein each column represents one of said themes and each row represents one of said themes; and

generating a score for at least a subset of said entries of said matrix, such that a score reflects a relative theme strength between two themes represented by said entry for said documents.

3.   The method as set forth in claim 2, wherein:

the step of extracting a plurality of themes further comprises the step of generating theme strengths for each theme extracted; and

the step of generating a score for at least a subset of said entries of said matrix comprises the steps of:

calculating a plurality of products for an entry by multiplying theme strengths corresponding to two themes represented by said entry for each document that includes said two themes represented by said entry; and

summing said products for an entry to generate said score.

4.   The method as set forth in claim 1, wherein the step of selecting category pairs in said knowledge base comprises the steps of:

determining whether only one of said themes exist as a category in said knowledge base;

if so,

generating a new category in said knowledge base for said theme;

generating a new cross-reference relationship between said new category and a category for which one of said themes exist; and

generating a new score for said new cross-reference relationship.

5.   The method as set forth in claim 1, wherein the step of selecting category pairs in said knowledge base comprises the steps of:

determining whether both of said themes exist as categories in said knowledge base;

if so,

determining whether a cross reference relationship exists from said category pair;

if not,

generating a new cross-reference relationship between said category pair;

generating a new score for said new cross-reference relationship; and

if so,

generating a new score for said existing cross-reference relationship.

6.   A system comprising:

search and retrieval module for receiving a user query and for generating a query response including query feedback;

a knowledge base, coupled to said search and retrieval module, for storing relationships among terminology for use as query feedback;

a knowledge base processing system, coupled to said knowledge base for processing a plurality of documents and automatically extending said relationships among said terminology in said knowledge base, said knowledge base processing system for extracting, from said documents, a plurality of themes, wherein a theme identifies subject matter contained in a corresponding document, for generating a plurality of scores such that each score identifies a relative theme strength among theme pairs of said themes extracted from said documents, said theme strength reflects the amount of subject matter contained in a document for a corresponding theme relative to other themes in said document, for selecting theme pairs based on said scores, for selecting category pairs in said knowledge base by mapping said themes of said theme pairs selected to corresponding categories of said knowledge base, and for generating a cross reference in said knowledge base between categories of said category pairs, wherein said cross reference identifies an association between said category pairs.

7.   The system as set forth in claim 6, wherein the knowledge base processing system further for generating a matrix comprising a plurality of columns and rows to form a plurality of entries, wherein each column represents one of said themes and each row represents one of said themes and for generating a score for at least a subset of said entries of said matrix, such that a

-69-

score reflects a relative theme strength between two themes represented by said entry for said documents.

8.   The system as set forth in claim 7, wherein the knowledge base processing system further for generating theme strengths for each theme extracted for calculating a plurality of products for an entry by multiplying theme strengths corresponding to two themes represented by said entry for each document that includes said two themes represented by said entry, and for summing said products for an entry to generate said score.

9.   The system as set forth in claim 7, wherein the knowledge base processing system further for determining whether only one of said themes exist as a category in said knowledge base, if so, for generating a new category in said knowledge base for said theme, for generating a new cross-reference relationship between said new category and a category for which one of said themes exist, and for generating a new score for said new cross-reference relationship.

10.   The system as set forth in claim 7, wherein the knowledge base processing system further for determining whether both of said themes exist as categories in said knowledge base; if so, for determining whether a cross reference relationship exists from said category pair; if not, for generating a new cross-reference relationship between said category pair, for generating a new

score for said new cross-reference relationship; and if so, for generating a new score for said existing cross-reference relationship.

11. A computer readable medium comprising a plurality of instructions, which when executed, causes the computer to perform the steps of:

extracting, from a plurality of documents, a plurality of themes, wherein a theme identifies subject matter contained in a corresponding document;

generating a plurality of scores such that each score identifies a relative theme strength among theme pairs of said themes extracted from said documents, said theme strength reflects the amount of subject matter contained in a document for a corresponding theme relative to other themes in said document;

selecting theme pairs based on said scores;

selecting category pairs in said knowledge base by mapping said themes of said theme pairs selected to corresponding categories of said knowledge base; and

generating a cross reference in said knowledge base between categories of said category pairs, wherein said cross reference identifies an association between said category pairs.

12. The computer readable medium as set forth in claim 11, wherein the step of generating a plurality of scores comprising the steps of:

generating a matrix comprising a plurality of columns and rows to form a plurality of entries, wherein each column represents one of said themes and each row represents one of said themes; and

generating a score for at least a subset of said entries of said matrix, such that a score reflects a relative theme strength between two themes represented by said entry for said documents.

13. The computer readable medium as set forth in claim 12, wherein:

the step of extracting a plurality of themes further comprises the step of generating theme strengths for each theme extracted; and

the step of generating a score for at least a subset of said entries of said matrix comprises the steps of:

calculating a plurality of products for an entry by multiplying theme strengths corresponding to two themes represented by said entry for each document that includes said two themes represented by said entry; and

summing said products for an entry to generate said score.

14. The computer readable medium as set forth in claim 11, wherein the step of selecting category pairs in said knowledge base comprises the steps of:

determining whether only one of said themes exist as a category in said knowledge base;

if so,

generating a new category in said knowledge base for said theme;

generating a new cross-reference relationship between said new category and a category for which one of said themes exist; and

generating a new score for said new cross-reference relationship.

15. The computer readable medium as set forth in claim 11, wherein the step of selecting category pairs in said knowledge base comprises the steps of:

determining whether both of said themes exist as categories in said knowledge base;

if so,

determining whether a cross reference relationship exists from said category pair;

if not,

generating a new cross-reference relationship between said category pair;

generating a new score for said new cross-reference relationship; and

if so,

generating a new score for said existing cross-reference relationship.